CHAPTER 5

Further characterisation of the *intB* element in *D. nodosus* strain A198

5.1 Introduction

There are two copies, or part thereof, of the *intB* element in the genome of the virulent *D. nodosus* strain A198 (Bloomfield *et al.*, 1997b). One copy of the *intB* element is present adjacent to the *attR* site of *vap* region 3, and sequencing analyses to date have led to the identification of five potential open reading frames, designated *intB*, *regA gepA*, *gepB* and *gepC* (Table 5.1) (Bloomfield *et al.*, 1997b). The location of this genetic element in relation to the *vap* regions is shown in Figure 5.1.

The second copy of the *intB* element sequences are present next to the *attR* site of *vap* region 2 in strain A198, and contains a partial copy of the *intB* gene, encoding of only the first 164 aa from the amino-terminus of the 403 aa *intB* gene, and is called $intB_N$ (Figure 5.1) (Bloomfield *et al.*, 1997b; Whittle, 1994). The $intB_N$ gene is followed by *orf379*, which is not related to *regA* or *gepB-C*.

In benign *D. nodosus* strain C305, there are two copies of the *intB* gene, both of which are truncated. One of these copies of the *intB* gene is integrated adjacent to *pnpA* in the C305 genome (Figure 5.1), and is identical to $intB_N$ from *D. nodosus* strain A198 (Bloomfield, 1997; Shaw, 1997). This is followed by *orf379*. The other copy of *intB* is next to the *intC* element sequences and consists of only 100 bp from the middle of the *intB*

Table 5.1: Sequence analysis of the *intB* element[a] in *D. nodosus* strain A198

Gene	Co-ordinates 5'-3- (nt)	Size (aa)	% identity to nt or putative protein	Homologues/Description	Accession Number	P(*n*)
intB	389-1599	403	33.1/314aa	*Coli* retronphage phiR73 integrase [*selCtRNA*]	M64113	7e-31
			32.8/323aa	*E. coli* 0157:H7 LEE pathogenicity island integrase [*tRNAsec*]	AF071034	6.9e-34
			32.7/315aa	*Mesorhizobium loti* symbiosis island integrase [*tRNAphe*]	AF049242	2.1e-27
			28.6/402aa	*Vibrio cholerae* 0395 integrase from CTXϕ PAI [*ssrA*]	U02372	4.3e-27
			33.6/303aa	*E. coli* K12 cryptic prophage P4-57 integrase [*ssrA*]	U11296	2.4e-26
			29.4/306aa	*Shigella flexneri* ϕSF6 integrase	X59553	2.7e-26
			29.7/306aa	*E. coli* K12 integrase [*tRNAargW*]	U11296	1.2e-21
			27.7/419aa	Bacteriophage P4 integrase	X05947	1.8e-19
			28.0/396aa	*E. coli* K12 integrase [*tRNAleuX*]	U14003	5.7e-19
			26.5/226aa	*Salmonella typhimurium* LT2 Gifsy-1 prophage [*lepA*]	AF001386	1.1e-14
			35.6/132aa	*Enetrobacter aerogenes* integrase	AF039582	4.2e-12
			23.4/385aa	*Pseudomonas putida* integrase from the *clc* element [*tRNAgly*]	PPAJ4950	1.2e-10
			36.6/123aa	*Rhodobacter capsulatus* putative prophage integrase	U57682	3.2e-07
			48.8/33aa	*Salmonella eneteridis* integrase from IS*3*-like element	SEJ002209	1.4e-04
			25.0/104aa	Bacteriophage adh site-specific recombinase IntG	Z97974	2.7e-02
			29.2/72aa	Lambda Bacteriophage phi80 integrase	X04051	5.4e-02
			22.7/154aa	*Streptococcus pyogenes* phage T12 integrase [*tRNAser*]	U40453	2.3e-01
			24.5/167aa	*S. typhimurium* site-specific recombinase XerC	U92525	3e-01
regA	1793-2491	233	39.2/232aa	*Pseudomonas aeruginosa* genes for negative regulator of pyocin genes, PrtR	D12706	3e-33
			38.0/245aa	*E. carotovora* RgdA DNA binding protein	L32173	8.3e-32
			36.2/235aa	Bacteriophage phi80 early region, *cI* gene product	X13065	1.4e-08
			28.1/135aa	Bacteriophage D3112 unidentified orf from *P. aeruginosa*	X52258	3.4e-08
			51.8/56aa	Bacteriophage 434 *cI* gene product	Y00118	6e-08
			27.6/87aa	*Staphlococcus aureus* Bacteriophage phi PVL proviral protein encoded by *orf19*	AB009866	3.9e-03
			21.4/210aa	*Haemophilus influenzae* unidentified putative protein	U32825	1.4e-01
gepA	3526-3056	157	31.7/70aa	*Bacteroides thetaiotomicron* conjugative transposon *rteC* gene putative protein	L02419	3.0e-03
			22.0/127aa	*Mycoplasma bovis* PG-45 variable surface protein A (VspA)	L81118	9.4
gepB	3624-4220	199	64.9/188aa	*Pseudomonas denitrificans orf7* putative protein downstream of *cob* genes	M62866	9.8e-55
gepC	>4595-4798	>67	-	NSH[b]	-	-

a. *intB* element from *D. nodosus* strain A198 (GenBank accession number X98546).
b. NSH indicates that there is no significant homology to sequences in databases.
c. Where the site of integration is known for an integrase gene it is indicated in square brackets following the desciption of the Int homologue.

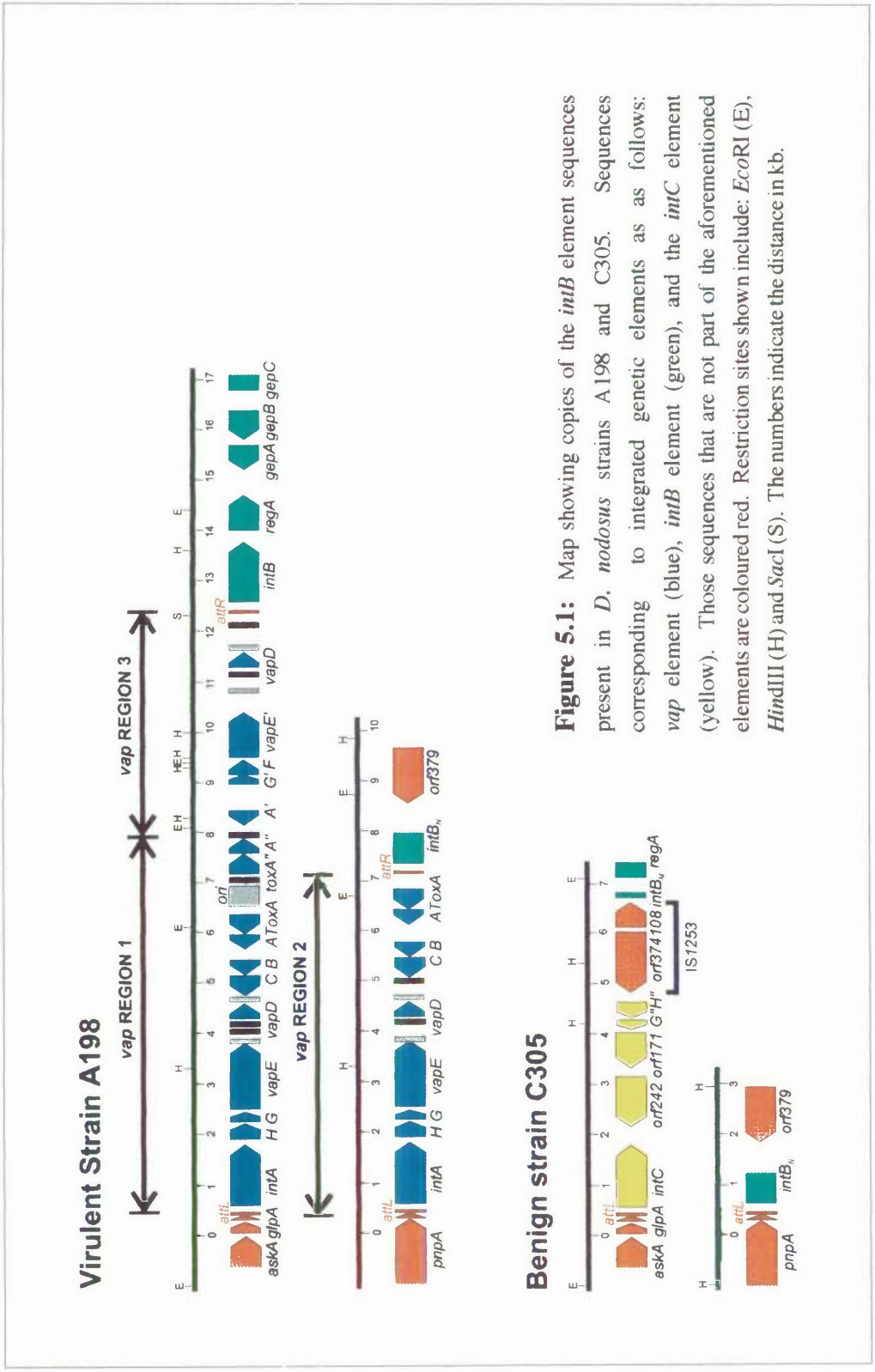


Figure 5.1: Map showing copies of the *intB* element sequences present in *D. nodosus* strains A198 and C305. Sequences corresponding to integrated genetic elements as follows: *vap* element (blue), *intB* element (green), and the *intC* element (yellow). Those sequences that are not part of the aforementioned elements are coloured red. Restriction sites shown include: *Eco*RI (E), *Hin*dIII (H) and *Sac*I (S). The numbers indicate the distance in kb.

coding region, and consequently is herein called $intB_M$ (Figure 5.1) (Bloomfield, 1997). $intB_M$ is followed by *regA*.

In summary, in strains A198 and C305 three variants of the *intB* gene have so far been sequenced. In an effort to further characterise the *intB* element, chromosome walking to the right of *gepC* in strain A198 was undertaken, and a sequence of 4.3 kb determined. The prevalence, arrangement and integrity of the *intB* element in seventeen strains of *D. nodosus* was investigated in Southern blot experiments, in order to determine if there is a correlation between the presence and integrity of the *intB* element, and virulence.

5.2 Results

5.2.1 Screening of *D. nodosus* strain A198 lambda library

A library of genomic DNA from the virulent *D. nodosus* strain A198 in bacteriophage lambda was screened using DNA fragments (Figure 5.2) derived from the *intB* gene (probe 1) and *gepBC* (probe 2) to probe filters from duplicate plaque lifts. Four lambda clones (λGW1.7, λGW1.5, λGW1.2 and λGW1.4) hybridising to both probes and a fifth clone hybridising to probe 2 alone (λGW1.6) were isolated (Figure 5.2). Restriction maps of overlapping lambda clones were constructed and aligned with previously-determined *intB* element sequences (Bloomfield *et al.*, 1997b) from strain A198 (Figure 5.2) which are adjacent to *vap* region 3 (Figure 5.1).

5.2.2 Sequencing of the region downstream of *gepC* in *D. nodosus* strain A198

DNA fragments from *Eco*RI, *Hin*dIII and *Eco*RI/*Hin*dIII restriction enzyme digests of λGW1.4 were gel purified and subcloned. In an effort to confirm that the sequences

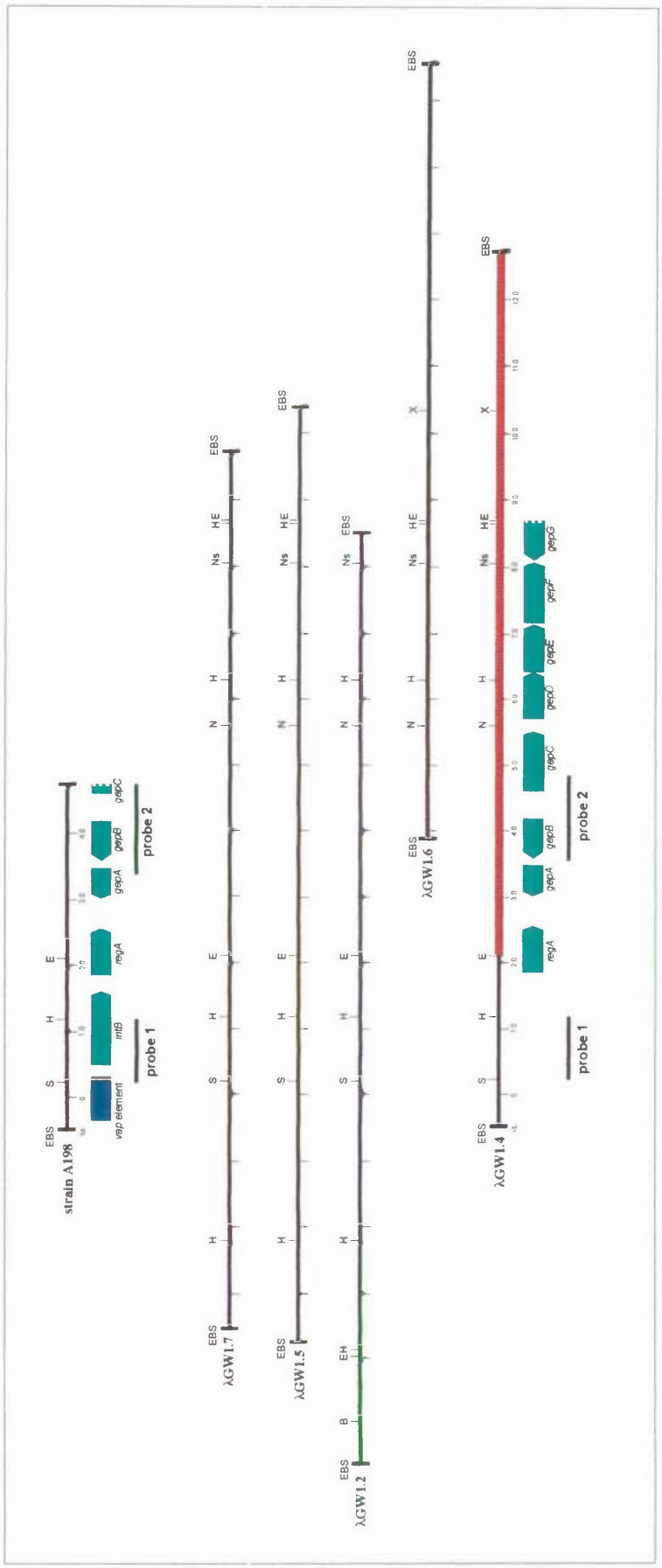


Figure 5.2: Restriction map of lambda clones isolated from a genomic library of *D. nodosus* strain A198 (Katz *et al.*, 1994), using probes specific for *intB* (probe 1) and *gepA-C* (probe 2). Restriction enzyme sites shown include: *Bam*HI (B), *Eco*RI (E), *Hin*dIII (H), *Nru*I (N), *Nsi*I (Ns) and *Sac*I (S). The numbers indicate the distance in kb. Sequences that were subcloned from λGW1.4 are indicated by a red box.

isolated on λGW1.4 reflected the arrangement of the same sequences in the A198 genome, a Southern blot analysis was performed. Genomic DNA from *D. nodosus* strain A198 was digested with *Eco*RI, *Hin*dIII and *Eco*RI/*Hin*dIII and probed with pGW39.1 (Figure 5.3). pGW39.1 hybridised to 7.0 kb *Eco*RI fragment, 5.3 kb and 2.5 kb *Hin*dIII fragments, and 4.4 kb and 2.5 kb *Hin*dIII/*Eco*RI fragments, which is consistent with the sizes predicted from the restriction enzyme map of λGW1.4 (Figure 5.3). A series of subclones extending from the *Eco*RI site at position 2114 to the most distal multiple cloning site in the right arm of bacteriophage lambda were constructed (Figure 5.3). A sequence of 4261 bp was determined (from 4671 nt right of the *Sac*I site to the *Eco*RI site at position 8932, Figures 5.3 and 5.4).

The *intB* element sequence (Figure 5.4) has a G + C content of 43%, which is similar to the 45% G + C content characteristic of the *D. nodosus* chromosome (Holdeman, Kelley & Moore, 1984). It was previously proposed that the *intB* element may have been acquired by horizontal transfer (Bloomfield *et al.*, 1997b), however, elements that have been acquired horizontally often have quite a different G + C content from the host chromosome. If the *intB* element was acquired horizontally, it may have been acquired long ago, and codon usage may have since evolved to be more like the host chromosome. Alternatively, it may have been acquired from an organism with a similar G + C composition.

The sequences upstream of *intB*, *regA* and *gepA-B*were analysed for open reading frames with start codons preceded by Shine-Dalgarno ribosome binding sites (Shine & Dalgarno, 1974), which led to the identification of five putative open reading frames, that have been designated genetic element protein genes *gepC*, *gepD*, *gepE*, *gepF* and *gepG* respectively in order to be consistent with earlier nomenclature (Bloomfield *et al.*, 1997b) (Figure 5.5).

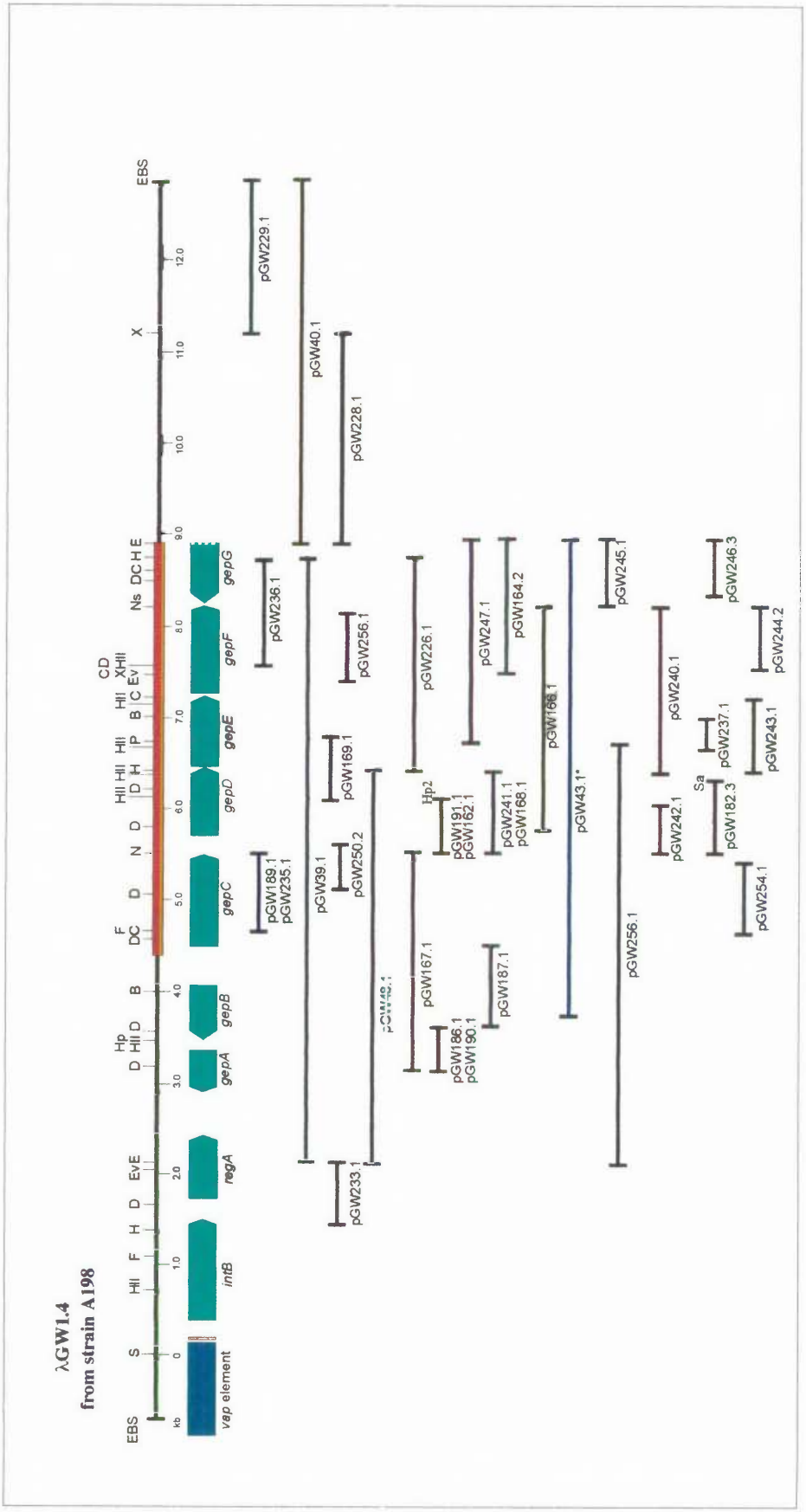
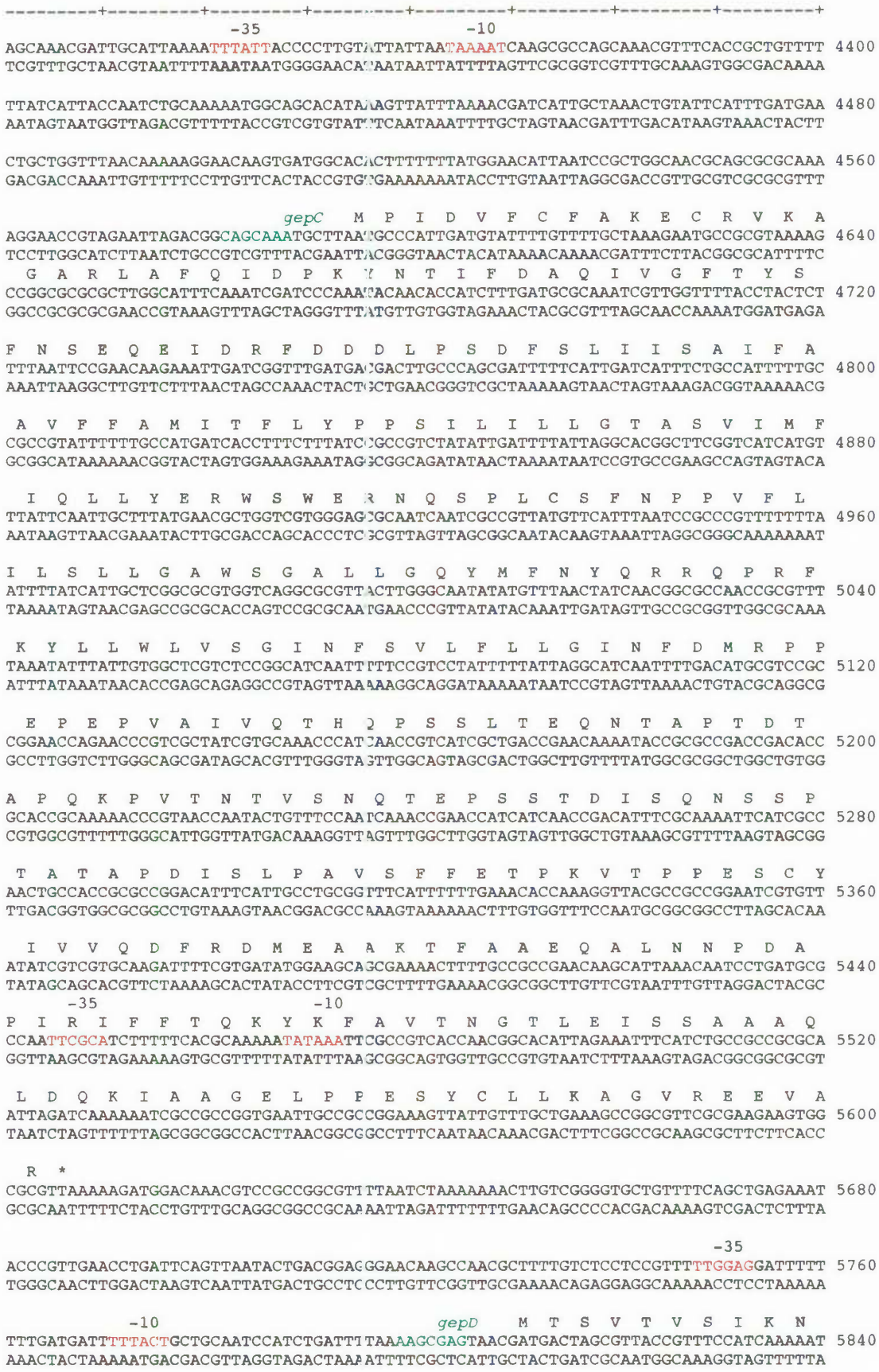
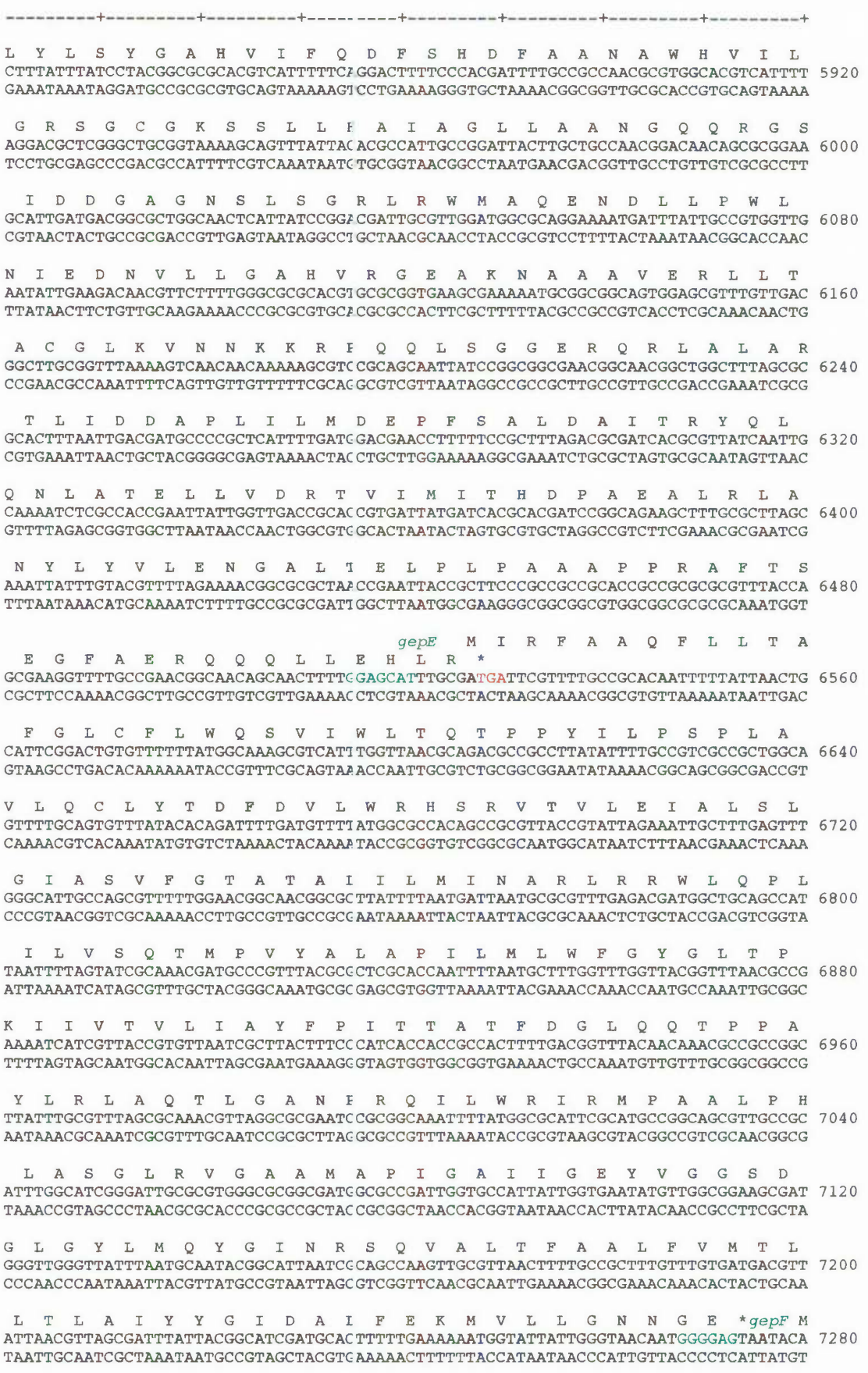


Figure 5.3: Restriction map of subclones from *D. nodosus* strain A198 lambda clone λGW1.4. Restriction sites are as follows: *Bbu*I (B), *Cla*I (C), *Dra*I (D), *Eco*RI (E), *Fsp*I (F), *Eco*RI (E), *Eco*RV (Ev), *Hin*dII (HII), *Hin*dIII (H), *Nru*I (N), *Nsi*I (Ns), *Sac*I (S) and *Xba*I (X). Restriction enzymes that recognise 4 bp recognition sequences are not shown on the map, but are instead shown at the ends of the subclones where applicable, as *Hpa*II (Hp2) and *Sau*3AI (Sa). Note that only restriction sites used in cloning are shown upstream of the *Eco*RI site at position 8925. The red box indicates the double stranded DNA sequence determined in this work. An asterisk marks a subclone which was derived from λGW1.6 DNA (Figure 5.3). The numbers indicate the distance in kb.


-35
-10
gepC M P I D V F C F A K E C R V K A
G A R L A F Q I D P K Y N T I F D A Q I V G F T Y S
F N S E Q E I D R F D D D L P S D F S L I I S A I F A
A V F F A M I T F L Y P P S I L I L L G T A S V I M F
I Q L L Y E R W S W E R N Q S P L C S F N P P V F L
I L S L L G A W S G A L L G Q Y M F N Y Q R R Q P R F
K Y L L W L V S G I N F S V L F L L G I N F D M R P P
E P E P V A I V Q T H Q P S S L T E Q N T A P T D T
A P Q K P V T N T V S N Q T E P S S T D I S Q N S S P
T A T A P D I S L P A V S F F E T P K V T P P E S C Y
I V V Q D F R D M E A A K T F A A E Q A L N N P D A
-35
-10
P I R I F F T Q K Y K F A V T N G T L E I S S A A A Q
L D Q K I A A G E L P P E S Y C L L K A G V R E E V A
R *
-35
-10
gepD M T S V T V S I K N


L Y L S Y G A H V I F Q D F S H D F A A N A W H V I L
CTTTATTTATCCTACGGCGCGCACGTCATTTTTCAGGACTTTTCCCACGATTTTGCCGCCAACGCGTGGCACGTCATTTT 5920
GAAATAAATAGGATGCCGCGCGTGCAGTAAAAAGTCCTGAAAAGGGTGCTAAAACGGCGGTTGCGCACCGTGCAGTAAAA
G R S G C G K S S L L A I A G L L A A N G Q Q R G S
AGGACGCTCGGGCTGCGGTAAAAGCAGTTTATTACACGCCATTGCCGGATTACTTGCTGCCAACGGACAACAGCGCGGAA 6000
TCCTGCGAGCCCGACGCCATTTTCGTCAAATAATGTGCGGTAACGGCCTAATGAACGACGGTTGCCTGTTGTCGCGCCTT
I D D G A G N S L S G R L R W M A Q E N D L L P W L
GCATTGATGACGGCGCTGGCAACTCATTATCCGGACGATTGCGTTGGATGGCGCAGGAAAATGATTTATTGCCGTGGTTG 6080
CGTAACTACTGCCGCGACCGTTGAGTAATAGGCCTGCTAACGCAACCTACCGCGTCCTTTTACTAAATAACGGCACCAAC
N I E D N V L L G A H V R G E A K N A A A V E R L L T
AATATTGAAGACAACGTTCTTTTGGGCGCGCACGTGCGCGGTGAAGCGAAAAATGCGGCGGCAGTGGAGCGTTTGTTGAC 6160
TTATAACTTCTGTTGCAAGAAAACCCGCGCGTGCACGCGCCACTTCGCTTTTTACGCCGCCGTCACCTCGCAAACAACTG
A C G L K V N N K K R Q Q L S G G E R Q R L A L A R
GGCTTGCGGTTTAAAAGTCAACAACAAAAAGCGTCCGCAGCAATTATCCGGCGGCGAACGGCAACGGCTGGCTTTAGCGC 6240
CCGAACGCCAAATTTTCAGTTGTTGTTTTTCGCACGCGTCGTTAATAGGCCGCCGCTTGCCGTTGCCGACCGAAATCGCG
T L I D D A P L I L M D E P F S A L D A I T R Y Q L
GCACTTTAATTGACGATGCCCCGCTCATTTTGATCGACGAACCTTTTTCCGCTTTAGACGCGATCACGCGTTATCAATTG 6320
CGTGAAATTAACTGCTACGGGGCGAGTAAAACTACCTGCTTGGAAAAAGGCGAAATCTGCGCTAGTGCGCAATAGTTAAC
Q N L A T E L L V D R T V I M I T H D P A E A L R L A
CAAAATCTCGCCACCGAATTATTGGTTGACCGCACCGTGATTATGATCACGCACGATCCGGCAGAAGCTTTGCGCTTAGC 6400
GTTTTAGAGCGGTGGCTTAATAACCAACTGGCGTCGCACTAATACTAGTGCGTGCTAGGCCGTCTTCGAAACGCGAATCG
N Y L Y V L E N G A L T E L P L P A A A P P R A F T S
AAATTATTTGTACGTTTTAGAAAACGGCGCGCTAACCGAATTACCGCTTCCCGCCGCCGCACCGCCGCGCGCGTTTACCA 6480
TTTAATAAACATGCAAAATCTTTTGCCGCGCGATTGGCTTAATGGCGAAGGGCGGCGGCGTGGCGGCGCGCGCAAATGGT
gepE M I R F A A Q F L L T A
E G F A E R Q Q Q L L E H L R *
GCGAAGGTTTTGCCGAACGGCAACAGCAACTTTTCGAGCATTTGCGATGATTCGTTTTGCCGCACAATTTTTATTAACTG 6560
CGCTTCCAAAACGGCTTGCCGTTGTCGTTGAAAACCTCGTAAACGCTACTAAGCAAAACGGCGTGTTAAAAATAATTGAC
F G L C F L W Q S V I W L T Q T P P Y I L P S P L A
CATTCGGACTGTGTTTTTTATGGCAAAGCGTCATTTGGTTAACGCAGACGCCGCCTTATATTTTGCCGTCGCCGCTGGCA 6640
GTAAGCCTGACACAAAAAATACCGTTTCGCAGTAAACCAATTGCGTCTGCGGCGGAATATAAAACGGCAGCGGCGACCGT
V L Q C L Y T D F D V L W R H S R V T V L E I A L S L
GTTTTGCAGTGTTTATACACAGATTTTGATGTTTTATGGCGCCACAGCCGCGTTACCGTATTAGAAATTGCTTTGAGTTT 6720
CAAAACGTCACAAATATGTGTCTAAAACTACAAAATACCGCGGTGTCGGCGCAATGGCATAATCTTTAACGAAACTCAAA
G I A S V F G T A T A I I L M I N A R L R R W L Q P L
GGGCATTGCCAGCGTTTTTGGAACGGCAACGGCGCTTATTTTAATGATTAATGCGCGTTTGAGACGATGGCTGCAGCCAT 6800
CCCGTAACGGTCGCAAAAACCTTGCCGTTGCCGCGAATAAAATTACTAATTACGCGCAAACTCTGCTACCGACGTCGGTA
I L V S Q T M P V Y A L A P I L M L W F G Y G L T P
TAATTTTAGTATCGCAAACGATGCCCGTTTACGCGCTCGCACCAATTTTAATGCTTTGGTTTGGTTACGGTTTAACGCCG 6880
ATTAAAATCATAGCGTTTGCTACGGGCAAATGCGCGAGCGTGGTTAAAATTACGAAACCAAACCAATGCCAAATTGCGGC
K I I V T V L I A Y F P I T T A T F D G L Q Q T P P A
AAAATCATCGTTACCGTGTTAATCGCTTACTTTCCCATCACCACCGCCACTTTTGACGGTTTACAACAAACGCCGCCGGC 6960
TTTTAGTAGCAATGGCACAATTAGCGAATGAAAGCGTAGTGGTGGCGGTGAAAACTGCCAAATGTTGTTTGCGGCGGCCG
Y L R L A Q T L G A N F R Q I L W R I R M P A A L P H
TTATTTGCGTTTAGCGCAAACGTTAGGCGCGAATCCGCGGCAAATTTTATGGCGCATTCGCATGCCGGCAGCGTTGCCGC 7040
AATAAACGCAAATCGCGTTTGCAATCCGCGCTTACGCGCCGTTTAAAATACCGCGTAAGCGTACGGCCGTCGCAACGGCG
L A S G L R V G A A M A P I G A I I G E Y V G G S D
ATTTGGCATCGGGATTGCGCGTGGGCGCGGCGATCGCGCCGATTGGTGCCATTATTGGTGAATATGTTGGCGGAAGCGAT 7120
TAAACCGTAGCCCTAACGCGCACCCGCGCCGCTACCGCGGCTAACCACGGTAATAACCACTTATACAACCGCCTTCGCTA
G L G Y L M Q Y G I N R S Q V A L T F A A L F V M T L
GGGTTGGGTTATTTAATGCAATACGGCATTAATCGCAGCCAAGTTGCGTTAACTTTTGCCGCTTTGTTTGTGATGACGTT 7200
CCCAACCCAATAAATTACGTTATGCCGTAATTAGCGTCGGTTCAACGCAATTGAAAACGGCGAAACAAACACTACTGCAA
L T L A I Y Y G I D A I F E K M V L L G N N G E *gepF M
ATTAACGTTAGCGATTTATTACGGCATCGATGCACTTTTTGAAAAAATGGTATTATTGGGTAACAATGGGGAGTAATACA 7280
TAATTGCAATCGCTAAATAATGCCGTAGCTACGTGAAAAACTTTTTTACCATAATAACCCATTGTTACCCCTCATTATGT

```
---------+---------+---------+---------+---------+---------+---------+---------+

K  K  I  S  T  F  L  F  G  L  M  L  A  T  T  A  L  A  K  E  P  L  H  L  M  L
TGAAAAAAATCAGTACGTTTTTATTCGGATTAATCTTGGCAACAACGGCTTTGGCAAAAGAGCCGCTGCATTTAATGTTG 7360
ACTTTTTTTAGTCATGCAAAAATAAGCCTAATTACAACCGTTGTTGCCGAAACCGTTTTCTCGGCGACGTAAATTACAAC

D  W  F  I  N  P  N  H  A  P  I  I  I  A  Q  Q  N  G  Y  F  D  K  H  G  L  E  V
GATTGGTTTATCAATCCCAATCATGCGCCGATCATCATCGCGCAACAAAACGGTTATTTTGATAAACACGGTTTAGAAGT 7440
CTAACCAAATAGTTAGGGTTAGTACGCGGCTAGTAGTAGCGCGTTGTTTTGCCAATAAAACTATTTGTGCCAAATCTTCA

 T  I  T  E  P  S  D  P  A  L  P  P  K  L  V  A  A  E  K  I  D  L  A  I  N  Y  Q
CACCATCACCGAACCGTCCGATCCGGCGCTGCCGCCGAAATTGGTTGCCGCGGAAAAAATCGATTTAGCCATCAATTATC 7520
GTGGTAGTGGCTTGGCAGGCTAGGCCGCGACGGCGGCTTTAACCAACGGCGCCTTTTTTAGCTAAATCGGTAGTTAATAG

  Q  Q  L  H  L  Q  I  D  E  G  L  P  I  S  R  V  S  T  L  I  A  T  P  L  N  C
AACAACAATTGCATTTACAAATTGACGAAGGATTCCCGATATCGCGCGTATCAACGTTAATCGCTACGCCTTTAAATTGC 7600
TTGTTGTTAACGTAAATGTTTAACTGCTTCCTAACGGCTATAGCGCGCATAGTTGCAATTAGCGATGCGGAAATTTAACG

V  I  V  D  A  Q  S  G  I  K  Q  V  S  D  L  K  G  K  K  I  G  Y  S  V  A  G  V
GTGATTGTTGACGCGCAAAGTGGCATTAAACAAGTTTCTGATTTAAAAGGTAAAAAAATTGGTTATTCCGTTGCCGGCGT 7680
CACTAACAACTGCGCGTTTCACCGTAATTTGTTCAAAGACTAAATTTTCCATTTTTTTAACCAATAAGGCAACGGCCGCA

 D  E  A  V  L  Q  S  F  L  A  S  G  G  L  T  L  N  D  V  K  L  V  N  V  N  F  S
TGATGAAGCCGTGTTGCAATCTTTTTTAGCCAGCCGTGGTTTAACTTTAAACGATGTGAAACTCGTCAACGTCAATTTTT 7760
ACTACTTCGGCACAACGTTAGAAAAAATCGGTCGCCACCAAATTGAAATTTGCTACACTTTGAGCAGTTGCAGTTAAAAA

  L  S  P  A  V  M  S  G  Q  V  D  A  V  I  G  A  A  R  T  V  E  L  H  E  M  K
CTTTATCGCCGGCGGTGATGAGTGGACAAGTAGACGCGGTTATCGGCGCCGCGCGCACGGTGGAATTACACGAAATGAAA 7840
GAAATAGCGGCCGCCACTACTCACCTGTTCATCTGCGCCAATAGCCGCGGCGCGCGTGCCACCTTAATGTGCTTTACTTT

A  H  N  H  E  G  R  A  F  F  L  E  E  H  G  I  P  P  F  D  E  L  I  F  V  A  H
GCGCACAATCATGAAGGGCGCGCCTTCTTTTTAGAAGAACACGGCATTCCGCCCTTTGATGAATTGATTTTTGTCGCGCA 7920
CGCGTGTTAGTACTTCCCGCGCGGAAGAAAAATCTTCTTGTGCCGTAAGGCGGGAAACTACTTAACTAAAAACAGCGCGT

 N  K  H  R  H  D  E  K  I  V  K  F  N  E  A  L  T  E  A  V  H  F  I  V  N  H  P
TAACAAACACCGCCACGATGAAAAAATCGTTAAATTTAATGAGGCGCTCACCGAAGCCGTGCATTTTATTGTCAATCACC 8000
ATTGTTTGTGGCGGTGCTACTTTTTTAGCAATTTAAATTACTCCGCGAGTGGCTTCGGCACGTAAAATAACAGTTAGTGG

  E  E  A  W  Q  K  Y  I  A  Y  K  K  G  L  D  D  A  V  N  Q  Q  A  W  K  D  S
CTGAAGAAGCATGGCAAAAATATATTGCTTATAAAAAAGGTTTAGATGATGCGGTTAATCAACAAGCATGGAAAGACAGT 8080
GACTTCTTCGTACCGTTTTTATATAACGAATATTTTTTCCAAATCTACTACGCCAATTAGTTGTTCGTACCTTTCTGTCA

L  T  R  F  A  L  R  P  A  A  L  D  D  R  R  Y  Q  N  Y  A  Q  Y  L  H  Q  I  G
TTAACGCGCTTTGCTTTGCGTCCGGCGGCGCTTGATGATCGGCGTTATCAAAATTACGCGCAATATTTGCATCAAATTGG 8160
AATTGCGCGAAACGAAACGCAGGCCGCCGCGAACTACTAGCCGCAATAGTTTTAATGCGCGTTATAAACGTAGTTTAACC

 L  I  K  K  I  V  P  V  S  E  Y  A  V  Q  P  *
TTTAATTAAAAAAATCGTGCCGGTTTCGGAGTATGCCGTGCAACCGTAATCGCGGTTAATCCTAAAAAGCCCTGTAACAG 8240
AAATTAATTTTTTTAGCACGGCCAAAGCCTCATACGGCACGTTGGCATTAGCGCCAATTAGGATTTTTCGGGACATTGTC

TGGCTTTTTTATGCATTTTATTGGCACGGGTTATAACGAAATGCCCGTGAGCTTTTTATATAAAGAAACGGGCGTAACTA 8320
ACCGAAAAAATACGTAAAATAACCGTGCCCAATATTGCTTTACGGGCACTCGAAAAATATATTTCTTTGCCCGCATTGAT
                                                  *  I  F  F  R  A  Y  S

TCCGTCATGCCGCTGACATAATCGCACACCGCCAATAATTTTTGATACGGGGAATCCGCATAACGACGATAACGCGTTGG 8400
AGGCAGTACGGCGACTGTATTAGCGTGTGGCGGTTATTAAAAACTATGCCCCTTAGGCGTATTGCTGCTATTGCGCAACC
 D  T  M  G  S  V  Y  D  C  V  A  L  L  K  Q  Y  P  S  D  A  Y  R  R  Y  R  T  P

CAATAATAATAAAATCATTTCATCGTGCCGGCGGCTGGTTTTATTGTTTAAAACCGTATTGGCAACCGCGGTGCAAAAAA 8480
GTTATTATTATTTTAGTAAAGTAGCACGGCCGCCGACCAAAATAACAAATTTTGGCATAACCGTTGGCGCCACGTTTTTT
  L  L  L  L  I  M  E  D  H  R  R  S  T  K  N  N  L  V  T  N  A  V  A  T  C  F

TATCCAATAATCCGCGCAAAATGCGGTTCCCCGCCACTTGAATTTCCACAACCGGTTGCAAGTTATAAATATAATCGCTG 8560
ATAGGTTATTAGGCGCGTTTTACGCCAAGGGGCGGTGAACTTAAAGGTGTTGGCCAACGTTCAATATTTATATTAGCGAC
I  D  L  L  G  R  L  I  R  N  G  A  V  Q  I  E  V  V  P  Q  L  N  Y  I  Y  D  S

GCAAATTGGCGCAACGCTTTTAATTGACGATAAGACGGAATATGCGGCAATAAAGCAGCGCCGTTTCCATTTAAAATTGC 8640
CGTTTAACCGCGTTGCGAAAATTAACTGCTATTCTGCCTTATACGCCGTTATTTCGTCGCGGCAAAGGTAAATTTTAACG
 A  F  Q  R  L  A  K  L  Q  R  Y  S  P  I  H  P  L  L  A  A  G  N  G  N  L  I  A

CGATTCTGATTGCCAAAATTGCGCGCACGTTTCATCGATTAATTGTCCAATCGCTTTGGCGCGCAAATAACCCAATTTAT 8720
GCTAAGACTAACGGTTTTAACGCGCGTGCAAAGTAGCTAATTAACAGGTTAGCGAAACCGCGCGTTTATTGGGTTAAATA
  S  E  S  Q  W  F  Q  A  C  T  E  D  I  L  Q  G  I  A  K  A  R  L  Y  G  L  K
```

```
---------+---------+---------+---------+---------+---------+---------+---------+

CGCTTTTACTGTGCATTGCTTTGATGCGCGTTTCATCAACATTTTCCCGCACCAATTCAAATAAACGCGCATAAGCTTCT 8800
GCGAAAATGACACGTAACGAAACTACGCGCAAAGTAGTTGTAAAAGGGCGTGGTTAAGTTTATTTGCGCGTATTCGAAGA
D  S  K  S  H  M  A  K  I  R  T  E  D  V  N  E  R  V  L  E  F  L  R  A  Y  A  E

TCAAAATGAATTTGTTTCACCTGATAAGCGTCTTCAATATCAACGATTAAATAACAAATATCATCAGCCGCTTCCATCAA 8880
AGTTTTACTTAAACAAAGTGGACTATTCGCAGAAGTTATAGTTGCTAATTTATTGTTTATAGTAGTCGGCGAAGGTAGTT
 E  F  H  I  Q  K  V  Q  Y  A  D  E  I  D  V  I  L  Y  C  I  D  D  A  A  E  M  L

ATACGATAACGGGTGCCGAAGCCAAACTTGCGGTTTAATTTGCGGAATTC 8930
TATGCTATTGCCCACGGCTTCGGTTTGAACGCCAAATTAAACGCCTTAAG
  Y  S  L  P  H  R  L  W  V  Q  P  K  I  Q  P  I  gepG
```

Figure 5.4: Nucleotide sequence of sequences upstream from *intB*, *regA* and *gepA-B* from *D. nodosus* strain A198. The amino acid sequences of five putative proteins encoded by *gepC*, *gepD*, *gepE*, *gepF* and *gepG* are aligned with the nucleotide sequence. Shine-Dalgarno sequences are indicated in green type, and putative –35 and –10 promoter sequences (red type).

5.2.3 Identification of a putative binding protein-dependent ABC transporter

The first 67 aa of the putative *gepC*-encoded protein had been determined previously (Bloomfield *et al*, 1997). In this work the remaining 270 aa of the putative GepC protein were determined. GepC has no similarity to sequences in the GenBank databases.

gepD encodes a potential protein of 239 aa which has a very high degree of amino acid identity to numerous ABC (ATP-binding Cassette) transporter, ATP-binding proteins (Table 5.2). *gepD* is in an operon-like arrangement with *gepE*, which overlaps the *gepD* coding region by one nucleotide. *gepE* encodes a protein which has very high similarity to ABC transporter membrane-associated proteins (Table 5.2). *gepF* encodes a protein that shares significant amino acid identity to Nmt1-like thiamine biosynthesis proteins (Table 5.2), which are, at least in *Haemophilus influenzae*, located immediately adjacent to genes encoding an ABC-transporter ATP-binding subunit and an ABC-transporter membrane-associated protein respectively (Table 5.2). *gepG* encodes a protein that shares similarity to hypothetical proteins of unknown functions.

Table 5.2: Sequence analysis of *gepC* to *gepG* from the *intB* element[a] in *D. nodosus* strain A198

Gene	Co-ordinates 5'-3- (nt)	Size (aa)	% identity to nt or putative protein	Homologues/Description	Accession Number	P(*n*)
gepC	4595-5606	337	-	NSH	-	-
gepD[b]	5811-6528	239	40.8/240	*Haemophilus influenzae* HI0354 ABC transporter, ATP-binding protein	U32720	9.2e-24
			38.1/218	*Escherichia coli* K12 MG1655 Orf255 ABC transporter, ATP-binding protein [taurine]	D85613	9.0e-16
			30.9/239	*Archaeglobus fulgidus* unidentified orf	U73857	4.3e-15
			34.2/242	*Pseudomonas putida* SsuB, ABC-type transporter, ATP-binding protein [sulfate]	AF075709	1.2e-14
			35.9/223	*Pseudomonas aeruginosa* AtsC, ABC-type transporter, ATP-binding protein [sulfate]	Z48540	1.6e-14
			36.1/202	*A. fulgidus* sulfate ABC transporter, ATP-binding protein AF0092 [sulfate]	AE001100	3.4e-14
			33.1/241	*Bacillus subtilis* YgaL hypothetical protein near nitrate transporter [nitrate]	Z93102	5.4e-14
			39.4/188	*B. subtilis* HisP importer [Histidine]	AF008220	6.2e-14
			34.5/246	*Mycobacterium* smegatis L5 and D29 bacteriophage resistance protein	U50335	8.2e-14
			30.1/229	*Serpulina hydrodysenteriae* ShiC putative ABC transporter [iron]	SHU75349	9.2e-14
			36.9/198	*Phormidium laminosum* (cyanobacterium) nrtC-Phl ATP-binding protein [nitrate]	Z19598	2.1e-13
			35.3/221	*Rhodobacter capsulatus* SB1003 ABC transporter, ATP-binding protein	AF010496	2.3e-13
			33.6/220	*H. influenzae* HitC iron utilisation protein [iron]	S72674	3.7e-13
			33.1/241	*B. subtilis* YglA, homologous to nitrate ABC transporter, ATP-binding protein [nitrate]	Z99108	4.7e-13
			34.6/205	*Synechococcus* sp. *ntrD* encoded nitrate transporter [nitrate]	X74597	8.7e-13
			36.4/217	*Methanococcus jannaschii* MJ0409 hypothetical protein	U67490	1.1e-12
			35.6/199	*Helicobacter pylori* HP0818 osmoprotection protein	AE000593	2.1e-12
			32.4/219	*Aspergillus nidulans* sulfate permease encoded by *cysA* [sulfate]	J04512	2.7e-12
			32.7/219	*Bacillus firmus* NatC ABC transporter, ATP-binding protein	AF084104	2.8e-12
			32.2/214	*Streptomyces coelicolor* GabT aminotransferase	AL031225	2.9e-12
			35.3/201	*Klebsiella pneumoniae* nitrate NasD nitrate transporter [nitrate]	KPNNASFEC	9.9e-12
			34.5/194	*H. influenzae* Rd thiamine ABC transporter, ATP-binding protein [thiamine]	U32782	1.0e-11
			31.2/231	*Salmonella typhiumrium* ATPase of 2-aminoethylphosphonate transporter	U69493	1.5e-11
			30.2/248	*Agrobacterium tumefaciens* pTi15955 MotB ATP-mannopine transport protein	U60011	1.7e-11
			32.3/211	*Streptococcus mutans* MsmK mannose transport protein [mannose]	M77351	2.3e-11

Table 5.2 continued: Sequence analysis of *gepC* to *gepG* of the *intB* element[a] in *D. nodosus* strain A198

gepE[b]	6527-7274	249	47.2/234	*H. influenzae* HI0355 ABC transporter permease protein [unknown]	U32720	2.8e-49
			30.8/227	*Chlamydia trachomatis* ABC transporter permease CT854 [unknown]	AE001358	4.4e-23
			27.8/234	*E. coli* K12 MG1655 TauC [taurine]	AE000143	1.2e-21
			25.1/211	*B. subtilis* YgaM, homolgous to nitrate permease protein [nitrate]	Z93102	1.9e-17
			26.9/234	*P. aeruginosa atsB* encoded ABC-type transporter, membrane subunit [sulfate]	Z48540	5.3e-16
			27.5/218	*M. jannaschii* MJ0409 hypothetical protein	U67490	1.8e-15
			27.1/229	*A. fulgidus* sulfate ABC transporter permease protein AF0093 [sulfate]	AE001100	7.7e-15
			26.8/231	*E. coli* K12 MG1655 OmpF outer membrane protein F precursor	AE000195	7.0e-14
			28.7/181	*E. coli* K12 ProW from osomoregulation operon	M24856	2.4e-13
			28.7/168	*Helicobacter pylori* HP0818 osmoprotection protein	AE000593	3.7e-13
			24.2/210	*P.aeruginosa* SsuC ABC transporter membrane subunit [sulfate]	AF075709	1.2e-12
			28.1/181	*E. coli* ProW glycine, betaine, proline transport protein	D90891	1.3e-12
			29.9/184	*B. subtilis* ProX osomoprotection protein	U38418	1.5e-10
			24.5/179	*Synechococcus* CmpB, homologous to NrtB [nitrate]	D26358	3.6e-10
			26.7/168	*A. fulgidus* AF0989 hypothetical protein	AE001036	8.2e-10
			25.7/187	*Synechocystis* sp. PCC6803 sll1715 hypothetical protein	D90916	1.1e-09
			31.7/142	*B. subtilis* choline transporter OpuBD	AF008930	1.5e-09
			20.1/201	*Synechococcus* PCC7942 periplasmic substrate CynB [cyanates]	AF001333	1.2e-08
			19.9/216	*B. subtilis* choline transporter OpuAB transmembrane protein	U17292	5.8e-08
			30.4/148	*B. subtilis* YvdK hypothetical protein	Z99121	6.8e-08
			26.1/172	*M. smegatis* L5 and D29 bacteriophage resistance protein	U50335	8.6e-08
			21.1/189	*K. pneumoniae* NasE nitrate transporter [nitrate]	L27431	1.9e-07
			26.6/184	*E. coli dld* encoded lactate dehydrogenase	AE000302	6.4e-07
			27.6/134	*Mycobacterium tuberculosis* H37RV DppB peptide transporter	AL022121	7.0e-05
			21.1/232	*Methylbacterium extorquens abcB* putative ABC transporter subunit B	U72662	6.0e-04
gepF	7355-8207	309	51.6/304	*H. influenzae* HI0357 putative thiamine biosynthesis protein encoded by *nmt1*	U32720	3.5e-82
			46.9/213	*Ochrobactum antropi* hypotheical protein	AQ242226	1.4e-46
			32.9/191	*Uromyces fabae* Nmt1 thiamine biosynthesis protein	UFU8179	8.7e-17
			27.7/173	*Saccharomyces pombe* Nmt1 thiamine biosynthesis protein (chromosome III)	J05493	1.3e-16
			28.4/190	*Aspergillus parasiticus* Nmt1	U15196	5.1e-16
			29.7/175	*Saccharomyces cerevisiae* TH15 thiamine biosynthesis protein (chromosome XIV)	AL031579	4.3e-15
			29.7/175	*S. cerevisiae* Nmt1 thiamine biosynthesis protein (chromosome IV)	Z74292	4.8e-15

Table 5.2 continued: Sequence analysis of *gepC* to *gepG* of the *intB* element[a] in *D. nodosus* strain A198

gepF ***(cont.)***	7355-8207	309	29.1/175	*S. cerevisiae* OrfY (chromosome X)	Z49656	1.4e-14
			22.4/232	*C. trachomatis* fumC encoded fumarate hydratase	AE001358	8.2e-11
			25.0/272	*A. fulgidus* AF0091 encoded protein gene, adjacent to sulfate transporter [sulfate]	AE001100	1.1e-10
			23.6/245	*B. subtilis orfK* encoded hypothetical protein	L16808	1.4e-05
			23.6/245	*B. subtilis* YzeA, homolgous to nitrate transport protein precursor	Z93102	5.6e-05
			36.4/77	*P. aeruginosa atsA* encoded arylsulfatase	Z48540	6.6e-04
			28.9/149	*Yersinia pestis* plasmid pMT1 unidentified orf near ABC transporter	AF053947	1.0e-03
			20.3/256	*M. bryantii* copper response extracellular protein	U40213	3.3e-02
			27.5/149	*Synechococcus* PCC7942 periplasmic substrate binding protein CynA, NrtA-like	AF001333	9.5e-01
gepG	>8930-8297	>210	33.0/203	*Synechocystis* sp. PCC6803 s110189 hypothetical protein	D64002	2.1e-17
			25.5/161	*H. influenzae* HI1299 hypotheical protein	U32809	1.0
			26.8/119	*P. aeruginosa* unidentified orf upstream of cysB encoding a transcription factor	U95379	10

a. NSH indicates that there is no significant homology to sequences in databases;
b Note that only the top twenty five homologues of these proteins have been shown.

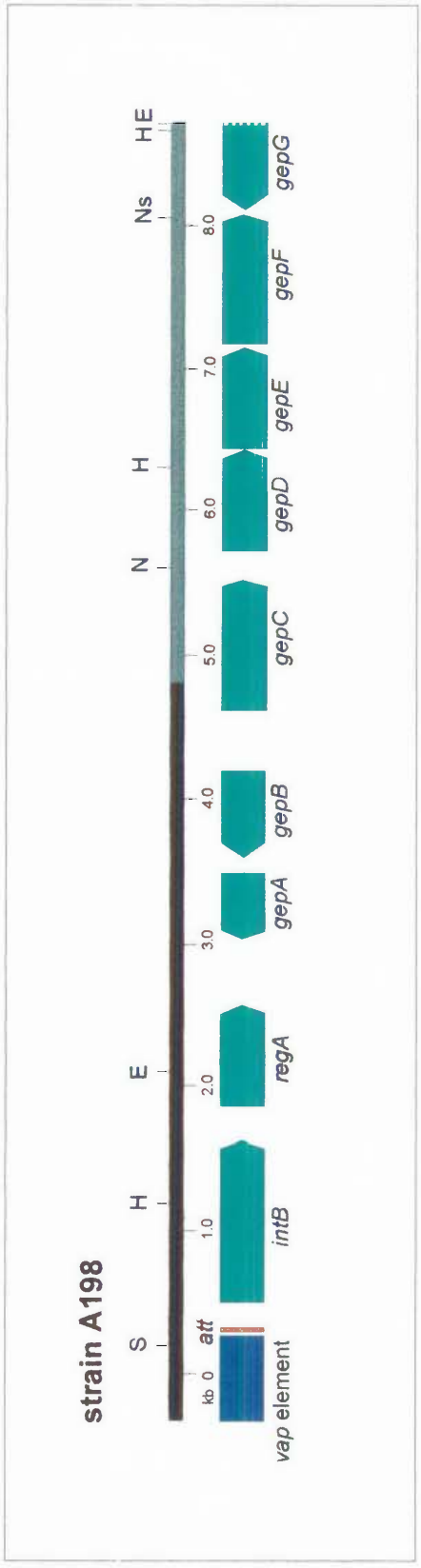


Figure 5.5: Restriction map of the *intB* element-associated sequences (green) from *D. nodosus* strain A198. The numbers shown indicate the distance in kb. Restriction sites shown include *Eco*RI (E), *Hin*dIII (H), *Nru*I (N), *Nsi*I (Ns) and *Sac*I (S). The sequences determined prior to this work are indicated by a black line, whilst those sequences determined in this work are indicated by a grey line above the putative orfs. The *vap* sequences which flank the *intB* element to the left are indicated by a blue box. A putative attachment site, *att*, is also shown (red rectangle).

It is interesting that the proteins to which the GepD and GepE putative transport proteins are most similar are involved in the import of substrates such as sulfate, nitrate, histidine, iron, thiamine, mannose and glycine/betaine (Table 5.2), and all belong to the binding protein-dependent (BPD) ABC transporter subfamily (Boos & Lucht, 1996).

These bacterial binding protein-dependent (BPD) ABC transporters belong to a superfamily of proteins called ABC transporters, which share a region of high homology which extends over a region of 200 aa (Boos & Lucht, 1996; Higgins, 1992) in one of the polypeptides. This 200 aa region is called the ATP-binding cassette, within which there is a highly conserved ATP-binding motif that contains two conserved sites (A and B) that form an ATP-binding pocket (Rossmann *et al.*, 1975).

This ATP-binding site typically occurs at the end of an α-helix, and the amino acid consensus sequence forms a turn which brings the lysine (K) residue in close proximity with one of the phosphate groups in the Mg^{2+}-ATP (Fath & Kolter, 1993). The negatively charged aspartate residue (D) interacts with the positively charged Mg^{2+} ion in the A-site (Fath & Kolter, 1993). The consensus sequence for the A-site has been defined as [AG]-x4-GK-[ST] (Fath & Kolter, 1993), whilst the B-site is defined as VLLLDEP (Boos & Lucht, 1996). These binding sites have been subsequently identified in bacterial proteins that are dedicated to the export and import of substrates and other proteins that are dependent on cellular energy (Fath, 1993; Higgins *et al.*, 1986).

The ATP-binding cassette also contains another conserved sequence of 9 bp called the linker peptide that is thought to play a critical role in signal transduction between the ABC subunit and the integral membrane protein domains (Boos & Lucht, 1996).

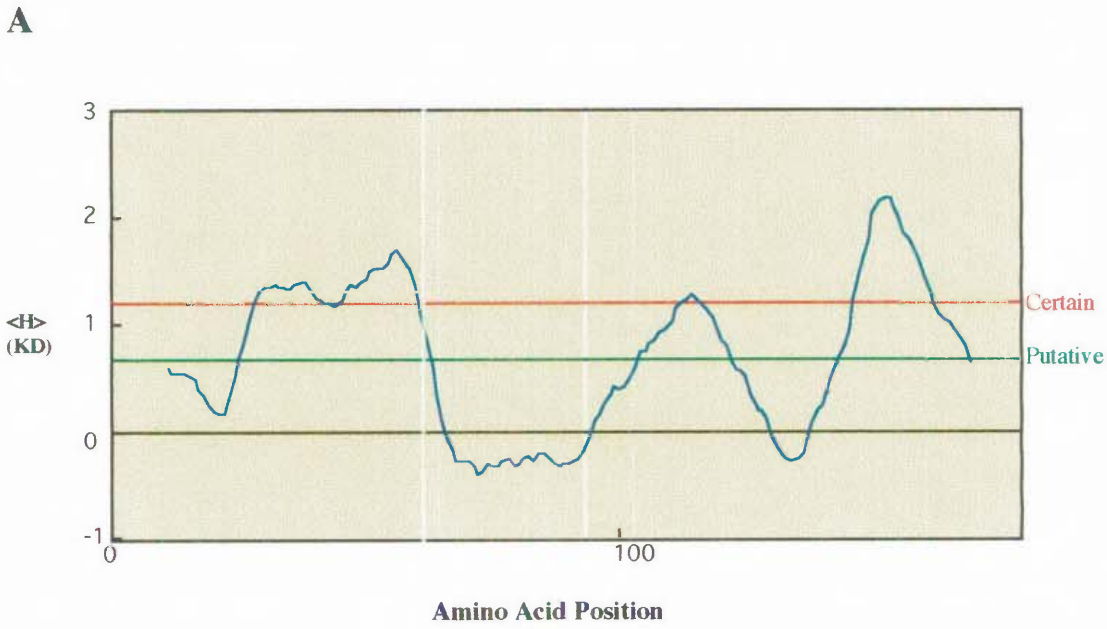
Using the WAG motif program and the Prosite patterns database, an ATP-binding cassette was identified within GepD, and was aligned with the ABC transporter consensus sequence as described in the literature (Boos & Lucht, 1996; Fath & Kolter, 1993)

(Figure 5.6). The ABC transporter, ATP-binding protein is not involved in substrate specificity (Boos & Lucht, 1996) but hydrolyses ATP and thereby provides the energy for the translocation of a wide variety of substrates across membranes (Higgins, 1992). The linker peptide motif is also present in GepD (Figure 5.6).

The membrane-associated protein (permease) component of BPD ABC transporters has a strongly hydrophobic character. Computer aided analysis (Claros, 1996; Claros & von Heijne, 1994) of hydrophobicity of the GepE protein, and predicted topology suggests that the GepE protein is an integral membrane protein that contains four putative transmembrane domains (Figure 5.7). Furthermore, these putative transmembrane domains are separated by hydrophilic peptide loops of varying length, two exposed to the periplasmic face and one to the cytoplasmic face. Both the amino and carboxyl termini are located on the cytoplasmic face of the membrane according to the "positive-inside rule" (von Heijne, 1986).

Most of the membrane-associated proteins of BPD transporters that have been described previously are predicted to have six transmembrane domains that form three periplasmic loops and two cytoplasmic loops, with both C- and N- termini on the cytoplasmic face (Higgins, 1992). However, few BPD ABC importers have been characterised (Boos & Lucht, 1996), and of those that have been characterised, some variation in the number of transmembrane domains has been noted, as determined by gene fusion and proteolysis experiments. For example, the ProW (514 aa) protein of *E. coli* contains seven transmembrane domains (Hennessey & Broomesmith, 1993), whilst HisQ (228 aa) and HisM (235 aa) of *S. typhimurium* are short for proteins in this class (~300 aa) and have only five transmembrane domains (Traxler, Boyd & Beckwith, 1993). The *D. nodosus* GepE (249 aa) protein is similarly short, and thus may not conform to the six transmembrane-spanner consensus structure characteristic of ABC transporters in general

Figure 5.6: ClustalW alignment of ABC transporter, ATP-binding proteins with similarity to the *gepD* encoded ATP-binding cassette (blue) and the consensus sequence for *E. coli* BPD ABC transporters (red) (Boos & Lucht, 1996) of the 200 aa domain containing the ABC. The ATP-binding motif domains (A) and (B) are underlined, and the linker peptide (L)is identified. The proteins aligned are from the following organisms: GepD (*D. nodosus*), HI0354 (*H. influenzae*), TauB (*E. coli*), AtsC (*P. aeruginosa*), AF0092 (*A. fulgidus*), TgaL (*B. subtilis*), ShiC (*S. hydrodystenterie*). Amino acids that are identical in the six sequences aligned with *gepD* (though in some cases different from the consensus) are indicated by asterisks; conserved amino acid residues are indicated by a period; 10 aa intervals are shown (v).



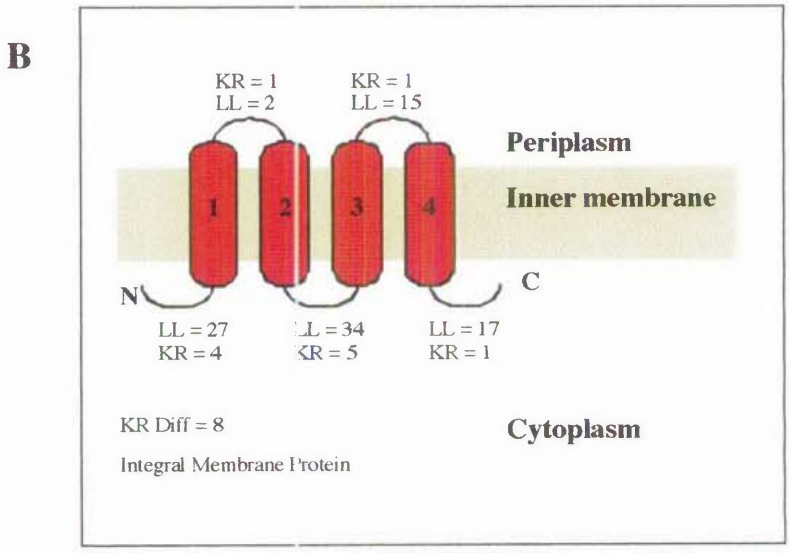


Figure 5.7: Hydrophobicity (<H>) plot (**A**) and predicted topology (**B**) of the GepE putative BPD ABC transporter membrane-associated protein using the computer program TopPred II (Claros, 1996). Hydrophilicity is calculated according to the algorithm of Kyte-Doolittle (KD) (Kyte & Doolittle, 1982). Transmembrane domains are numbered 1 to 4. Other features are as follows: LL indicates the number of amino acids in the loop; KR indicates the number of lysine and arginine residues respectively; KR Diff indicates the positive charge difference; N- and C- indicate the amino and carboxy termini of the protein. Cytoplasmic and periplasmic faces and the inner membrane are also indicated.

(Fath & Kolter, 1993; Higgins, 1992). Alternatively, the observation that GepE does not contain the expected number of transmembrane subunits may also be a limitation of the program used to predict the topology of the GepE protein.

Despite the high degree of similarity between the ATP-binding cassette motifs of different ABC transporter proteins in general, there is often little sequence similarity between the membrane-associated proteins of different ABC transporters. It has been proposed that the folding of a membrane protein is likely to be less sensitive to amino acid substitutions than a soluble protein, and hence hydrophobic domains tend to be more divergent than hydrophilic ones (Higgins, 1992).

A ClustalW alignment of the putative ABC transporter membrane-associated protein, GepE, with similar transmembrane proteins, shows that although all have similarity to GepE (Table 5.2), they are divergent from each other. Despite this divergence, there are seven amino acid residues that are absolutely conserved including three glycine, three proline and one phenylalanine (Figure 5.8). This suggests that these particular residues may have an important functional or structural role in these related proteins.

Glycine, is often found at turns in the protein secondary structure, primarily because where it is present in a protein the minimal steric hindrance of the glycine side chain results in more structural flexibility than other amino acids. Proline is the opposite of glycine, since the secondary amino group is held in a rigid conformation that decreases the flexibility of the protein at the point at which it is present (Lehninger, Nelson & Cox, 1993). The prediction of α-helices, β-sheets, and turns in GepE was done using the GCG peptidestructure program (Jameson & Wolf, 1988) (Figure 5.8) available through ANGIS. The position of conserved proline and glycine residues in GepE relative to potential

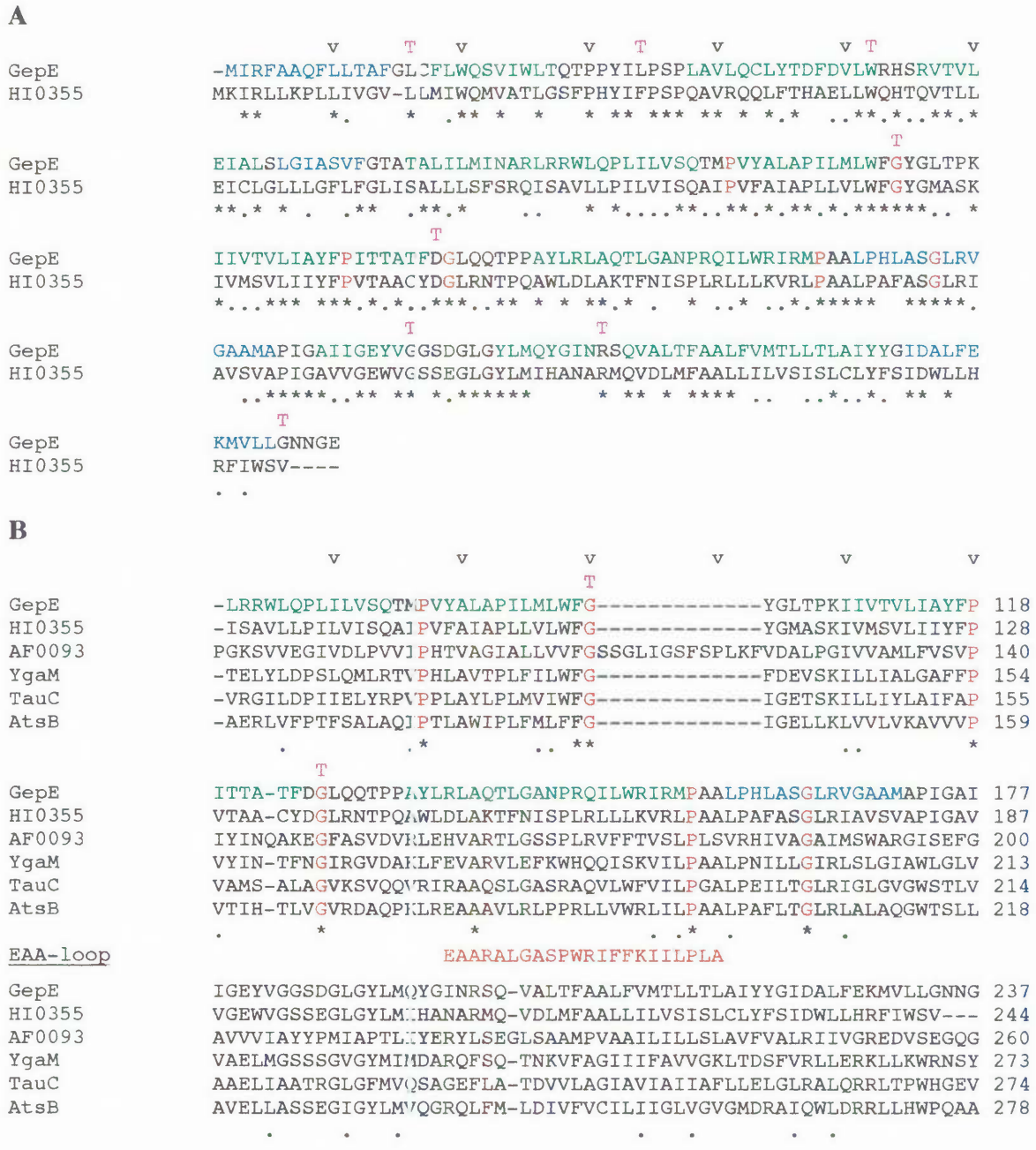


Figure 5.8: ClustalW alignment of ABC transporter membrane proteins with similarity to the *gepE* encoded protein. The amino acid alignment between highly related proteins GepE (*D. nodosus*) and HI0355 (*H. influenzae*) is shown (**A**). The GCG program peptidestructure was used to predict secondary structures in GepE according to the methods of Chou-Fasman and Garnier-Osguthorpe-Robson. Secondary structures are indicated as follows: blue (α-helix), green (β-sheet) and turns (T). (**B**) The amino acid alignment between GepE, HI0355 and more distantly related proteins AF0093 (*A. fulgidus*), YgaM (*B. subtilis*), TauC (*E. coli*) and AtsB (*P. aeruginosa*). The position in the amino acid sequences of the aligned proteins is indicated at the end of each line. Amino acids that are identical in all six sequences are indicated by asterisks; conserved amino acid residues are indicated by a period; 10 aa intervals are shown (v). The conserved glycine and proline residues are shown in red. The EAA-loop is shown in red.

secondary structures is consistent with the hypothesis that these residues may be important to the structural and therefore functional integrity of the protein.

In the membrane-associated protein of the BPD transport systems there is a single conserved sequence between 94 and 115 aa from the C-terminus feature which has the consensus sequence EAA---G----------I-LP (Dassa & Hofnung, 1985), called the EAA-loop. More recent studies of forty-seven membrane subunits led to the identification of a similar sequence in all proteins but with variations in some positions of this consensus sequence (Benner *et al.*, 1994). Similar variations are evident in the ClustalW alignment of GepE and related sequences (Figure 5.8). The exact function of the EAA-loop has not yet been elucidated, however it has been proposed that since it is located on the cytoplasmic side of the membrane it may interact with a ligand or protein, possibly the ATP-binding subunit of the BPD ABC transporter (Kerppola *et al.*, 1991).

5.2.4 A role for *gepC*?

In two-thirds of BPD transporters so far characterised, there are two separate membrane-associated proteins (Boos & Lucht, 1996), and it is assumed that these two membrane-associated proteins form a heterodimer within the transport complex. In the other third, the membrane-associated protein is thought to form a homodimer (Boos & Lucht, 1996; Fath & Kolter, 1993; Higgins, 1992).

In the *gep* operon only GepE had similarity to membrane-associated proteins of BPD importer proteins. However, computer aided analysis of the hydrophobicity and predicted topology of the GepC protein indicates that the GepC protein is also an integral membrane protein (Figure 5.9). The GepC has three potential transmembrane domains, that are separated by two hydrophilic loops, and an N- terminus which is located on the cytoplasmic face, whilst the C-terminus is found on the periplasmic side of the membrane.